Category-level Multi-Part Multi-Joint 3D Shape Assembly

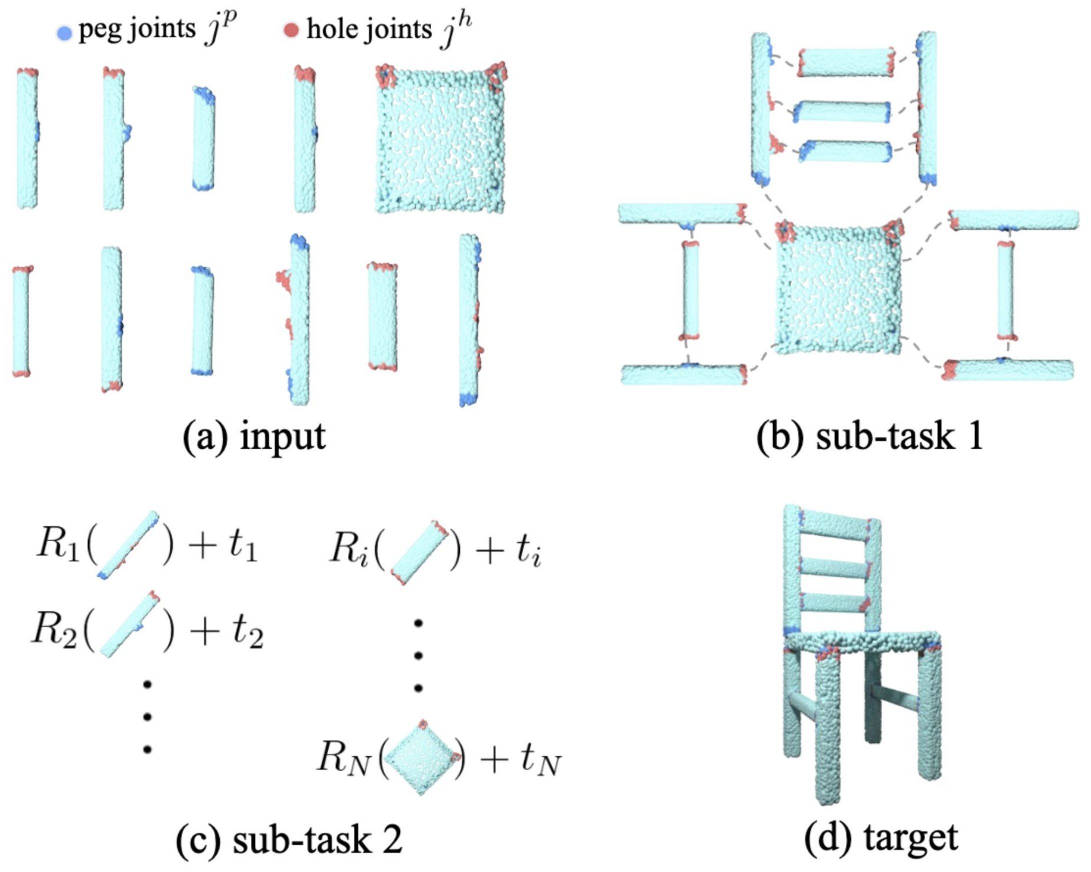
Yichen Li[1,2], Kaichun Mo[3], Yueqi Duan[4], He Wang [5], Jiequan Zhang[2], Lin Shao[6], Wojciech Matusik[1], Leonidas Guibas[2]

[1] MIT CSAIL [2] Stanford [3] Nvidia [4] THU [5] PKU [6] NUS

Observation & Motivation

Prior works on 3D shape assembly [1, 2] ignore the matching between joints, generating geometrically consistent shapes, but cannot be used for assembly in the real world.

We introduce the notion of joints and propose a new 3D shape assembly task from a multi-part multi-joint perspective. The tasks involves a bilateral optimization of discrete joint matching and continuous pose prediction.



Method

Shape Loss [1]

$$\mathcal{L}_t = \sum_{i=1}^{N} ||t_i - t_i^{gt}||_2^2,$$

$$\mathcal{L}_r = \sum_{i=1}^{N} d_{chamfer}(R_i(p_i), R_i^{gt}(p_i))$$

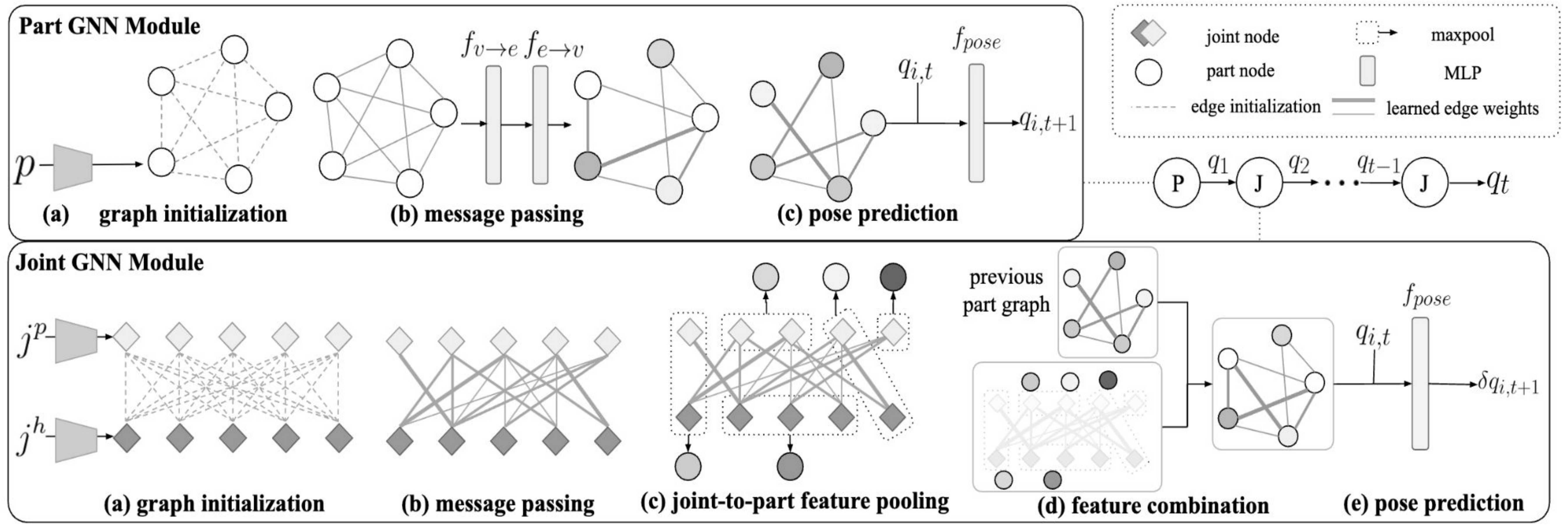
$$\mathcal{L}_a = d_{chamfer}\left(\sum_{i=1}^{N}(q_i(p_i)), \sum_{i=1}^{N}(q_i^{gt}(p_i))\right)$$

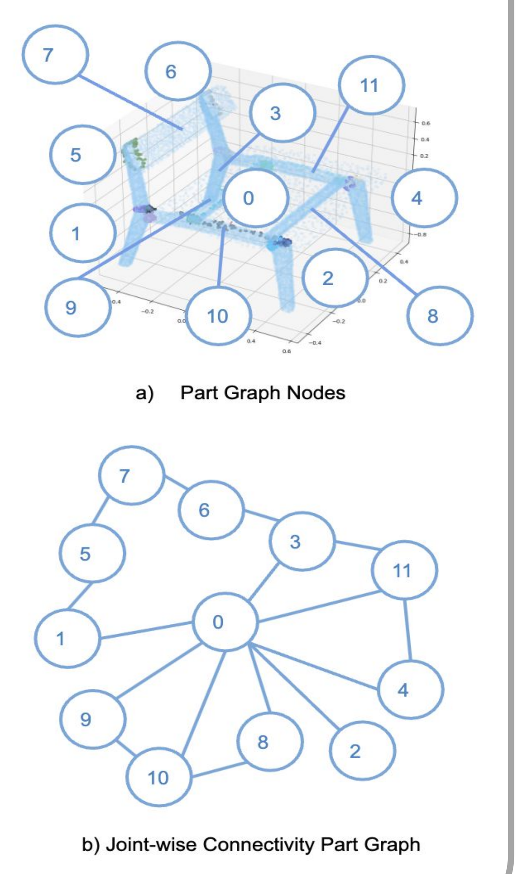
Joint Loss

$$\mathcal{L}_{flip} = \sum_{i=1}^{N} ||q_i(p_i) - q_i^{gt}(p_i)||_F^2$$

$$\mathcal{L}_{coarse} = \sum_{\phi_i=1}^{M} ||q_a(\bar{j}_a^p) - q_b(\bar{j}_b^h)||_2^2$$

$$\mathcal{L}_{fine} = d_{chamfer}(q_a(j_a^p), q_b(j_b^h)), \quad j_a^p \in p_a, j_b^h \in p_b$$





Results

Shape Metric [1]

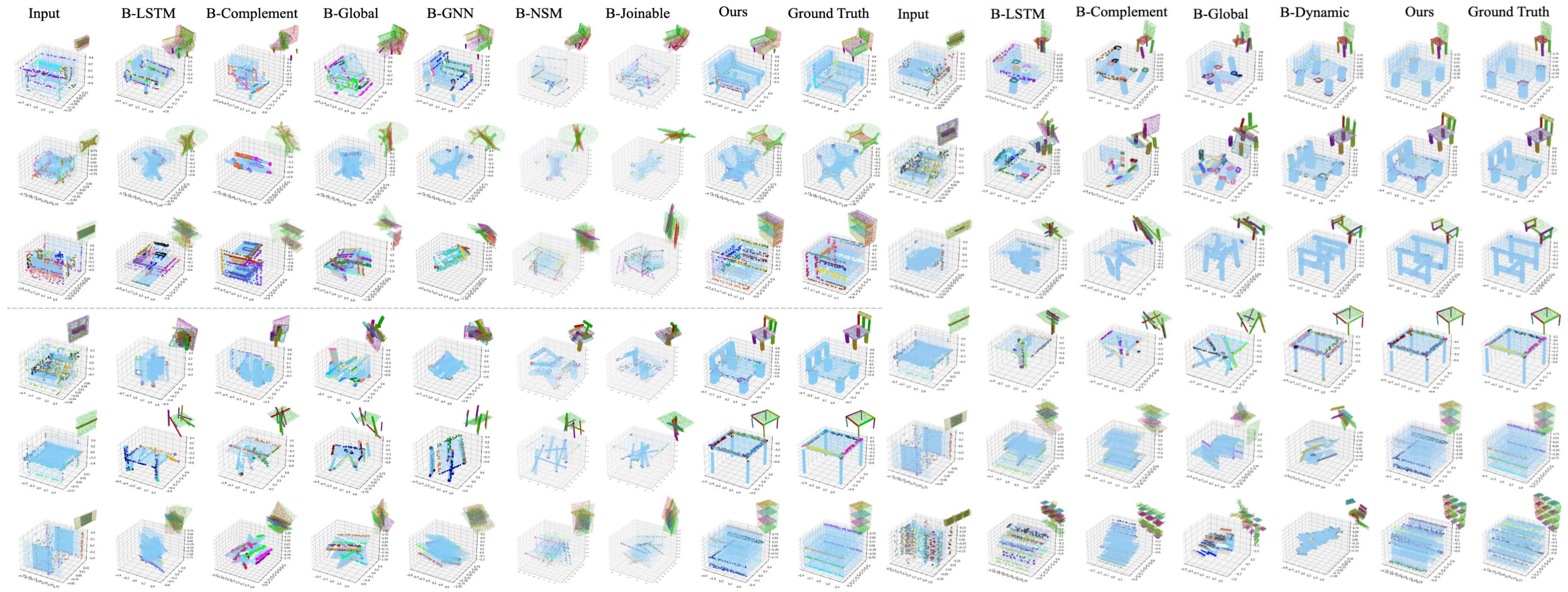
$$shape\ CD = \frac{1}{N \cdot n_{points}} \sum_{i=1}^{N} d_{chamfer}\left(q_i(p_i), q_i^{gt}(p_i)\right)$$

$$Part\ Acc. = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left[d_{chamfer}\left(q_i(p_i), q_i^{gt}(p_i)\right)\right] < \tau_p$$

Joint Metric

$$Joint\ CD = \frac{1}{M}\sum_{\phi_i=1}^{M} d_{chamfer}\left(q_a(j_a^p), q_b(j_b^h)\right)$$

$$Joint\ Acc. = \frac{1}{M}\sum_{\phi_i=1}^{M} \mathbb{1}\left[d_{chamfer}(q_b(j_b^h), q_a(j_a^p))\right] < \tau_j$$



		Shape Chamfer Distance ↓				Part Pose Accuracy ↑			
Setting	Method	Chair	Table	Cabinet	Average	Chair	Table	Cabinet	Average
Original Setting	B-Global	0.015	0.013	0.008	0.013	32.8	30.1	33.6	31.4
	B-LSTM	0.017	0.026	0.007	0.021	39.4	22.5	44.4	30.8
	B-Complement	0.028	0.034	0.222	0.046	11.0	5.33	0.0	7.2
	B-GNN	0.007	0.008	0.006	0.007	65.3	61.4	45.0	61.7
	B-NSM	0.013	0.022	0.012	0.018	25.3	48.2	18.9	37.0
Our Full Setting (joint input and loss)	B-Global	0.029	0.022	0.013	0.024	5.4	12.0	15.0	9.6
	B-LSTM	0.037	0.029	0.017	0.031	4.4	4.1	15.3	5.1
	B-Complement	0.048	0.044	0.029	0.044	4.5	8.0	11.6	6.9
	B-GNN	0.034	0.039	0.021	0.036	11.5	3.2	10.4	7.0
	B-NSM	0.014	0.032	0.020	0.024	19.0	12.1	14.7	15.0
	B-Joinable	0.026	0.037	0.025	0.032	12.6	7.3	12.1	9.7
	Ours	**0.006**	**0.007**	**0.005**	**0.006**	**72.8**	**67.4**	**63.3**	**69.2**

		Joint Chamfer Distance ↓				Joint Matching Accuracy ↑			
Setting	Method	Chair	Table	Cabinet	Average	Chair	Table	Cabinet	Average
Original Setting	B-Global	0.712	0.847	0.667	0.780	13.4	15.8	10.7	14.5
	B-LSTM	0.756	0.728	0.651	0.733	17.0	13.2	14.8	14.8
	B-Complement	0.901	0.977	1.074	0.954	7.5	8.3	23.6	9.2
	B-GNN	0.725	0.855	0.683	0.791	24.4	30.0	18.6	26.9
	B-NSM	0.697	0.717	0.700	0.708	15.1	16.9	17.1	16.2
Our Full Setting (joint input and loss)	B-Global	0.513	1.268	0.488	0.912	12.7	4.0	6.9	7.6
	B-LSTM	0.394	0.875	0.467	0.655	20.3	7.7	13.8	13.1
	B-Complement	0.456	0.647	0.503	0.561	17.2	15.5	17.0	16.3
	B-GNN	0.379	0.786	**0.416**	0.598	21.5	10.3	20.0	15.4
	B-NSM	0.556	0.698	0.517	0.629	18.9	12.1	7.8	14.4
	B-Joinable	0.653	0.812	0.483	0.725	16.1	13.9	9.4	14.4
	Ours	**0.352**	**0.602**	0.620	**0.505**	**57.2**	**50.6**	**27.5**	**51.4**

Impacts & Conclusions

Our work has the following contributions:

- We consider the concept of joint for the problem of category-level multi-part 3D shape assembly.
- We introduce a joint-annotated part dataset as well as a set of evaluation metrics to examine the performance.
- We propose a novel hierarchical graph network that simultaneously optimizes for both holistic shape structure and the joint alignment accuracy.
- Extensive experiments show that our method not only improves both shape structure and joint alignment over baseline method.

We note several possible future directions:

- Extend to upstream task of joint detection from part geometries.
- Extend to more complicated joints for this problem and construct a general formulation for all possible joint types.
- Sequential planning for joint alignment to better enable vision algorithms to be deployed in autonomous systems.

References

[1] Yichen Li, Kaichun Mo, Lin Shao, Minhyuk Sung, and Leonidas Guibas. Learning 3d part assembly from a sigle image. In Computer Vision–ECCV 2020.

[2] Yun-Chun Chen, Haoda Li, Dylan Turpin, Alec Jacobson, and Animesh Garg. Neural shape mating: Self-supervised object assembly with adversarial shape priors. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[3] Karl DD Willis, Pradeep Kumar Jayaraman, Hang Chu, Yunsheng Tian, Yifei Li, Daniele Grandi, Aditya Sanghi, Linh Tran, Joseph G Lambourne, Armando Solar-Lezama, and Wojciech Matusik. Joinable: Learning bottom-up assembly of parametric cad joints. In Proceedings of the IEEE/CVF CVPR 2022.